

Conference paper

Data management challenges in a distributed organization: What challenges we are facing at the USGS and how are we working to overcome them.

JC Nelson, U.S. Geological Survey

Abstract

In 2013, a series of White House orders required that federal data be made open and machine readable to the public. In response, the U.S. Geological Survey (USGS) developed a Public Access Plan and internal policies to meet these open data requirements. Every USGS Science Center is required to respond to and put in place measures to meet the guidelines as laid out in the Public Access Plan. The center plans not only have to meet the USGS plan, but also must ensure they continue to meet Mission Area goals and Regional priorities.

Background

In 2013, President Obama signed the Executive Order “Making Open and Machine Readable the New Default for Government Information” to make data collected by the government more open and machine-readable (Obama 2013). This along with White House memoranda released by the Office of Science and Technology Policy called “Increasing Access to the Results of Federally Funded Scientific Research” (Holdren et al. 2013) and the Office of Management and Budget called “Open Data Policy – Managing Information as an Asset” (Burwell et al. 2013) that supports transparency by addressing the dissemination and interoperability of data and information products produced by federal agencies.

In response, the U.S. Geological Survey (USGS) published the plan “Public Access to Results of Federally Funded Research at the U.S. Geological Survey”, which discusses the USGS’s unified approach to make federally funded data more available (USGS 2016a). Moreover, USGS has enacted a series of Instructional Memoranda (IM) which provide specific data management and data publishing policies (USGS 2015) for its staff to successfully meet the new open data requirements.

A Distributed Organization

Coordinating data management activities to comply with Federal guidelines can be challenging for any organization. However, the vast mission and distributed organizational structure of the USGS adds to this challenge. The USGS is represented by almost 10,000 scientists, technicians, and support staff working in more than 400 locations across the United States (USGS 2016b). USGS science is coordinated by 7 Mission Areas and 7 Regions.

The Mission Areas include:

- Climate and Land Use Change
- Core Science Systems
- Ecosystems
- Energy and Minerals

- Environmental Health
- Natural Hazards
- Water

The Regions include:

- Alaska
- Midwest
- Northeast
- Northwest
- Pacific
- Southeast
- Southwest

A majority of this presentation will be about the lessons learned within the Ecosystems Mission Area. There are 16 Ecosystems (ECO) Science Centers in 15 states, with numerous field stations throughout the rest of the country. This includes at least 1 ECO Science Center in each of the 7 USGS Regions.



Figure: The mission of the USGS Ecosystems Science Centers is as vast as its spatial representation of the United States. (<https://www2.usgs.gov/ecosystems/map.html>)

Challenges

There are many challenges associated with data management and with working in a distributed organization. The biggest challenge stemming from the USGS Public Access Plan (USGS 2016a) is that this mandate comes with no funding attached to its implementation. There is no additional funding provided for use in meeting the requirements. The centers also have to develop responses to the public access plan while still maintaining their overall missions of the center and the Mission Area. Additionally, the needs and

resources of the Region have to be addressed. The extensive mission of the USGS means that every center and program, have different needs and requirements making a size fits all solution very difficult to implement. The USGS also has a strong commitment to work with a wide range of partners that also have different data management elements that need to be met. In this presentation, examples of the organizational matrix will be used to demonstrate the complexities faced by science centers.

Science centers also possess varying levels of existing expertise related to data management and the Science Data Lifecycle (Faundeen et al, 2013). If skills do not exist at a center, they have to train staff or bring in those skills from outside of the center. The centers, regions, and the Mission Areas all have ongoing efforts to address these challenges.

Acknowledgements

The author would like to thank the USGS Community for Data Integration (CDI) and the CDI Data Management Working Group members for their ongoing contributions. The author would also like to thank the USGS's Ecosystems and Core Science Systems Mission Areas for funding and guidance. A special thanks also goes out to the USGS's Midwest Region for access and administrative support to 19 science centers representing 3 Mission Areas and again to the USGS's Ecosystem Mission Area for access to 16 science centers in 7 regions.

Competing Interests

The author declares that they have no competing interests.

References

Faundeen, J L, Burley, T E, Carlino, J A, Govoni, D L, Henkel, H S, Holl, S L, Hutchison, V B, Martín, E, Montgomery, E T, Ladino, C C, Tessler, S, and Zolly, L S 2013 The United States Geological Survey Science Data Lifecycle Model: U.S. Geological Survey Open-File Report 2013–1265, 4 p. DOI: <http://dx.doi.org/10.3133/ofr20131265>.

Holdren, J P 2013 Memorandum for the heads of executive departments and agencies—Increasing Access to the Results of Federally Funded Scientific Research: Executive Office of the President, Office of Science and Technology Policy (February 22, 2013), 6 p. Available at http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf [Last accessed 9 May 2016].

Obama, B 2013 Executive Order No. 13642, Making Open and Machine Readable the New Default for Government Information: Federal Register Vol. 48, No. 93, pg. 28111 (May 9, 2013), Available at <https://www.federalregister.gov/articles/2013/05/14/2013-11533/making-open-and-machine-readable-the-new-default-for-government-information> [Last accessed 9 May 2016].

U.S. Geological Survey 2015 Fundamental Science Practices Policies: Office of Science Quality and Integrity (OSQI) Instructional Memoranda (IM) - IM OSQI 2015-01, IM OSQI 2015-02, IM OSQI 2015-03, IM OSQI 2015-04. Available at <http://www2.usgs.gov/fsp/policies.asp> [Last accessed 9 May 2016].

U.S. Geological Survey 2016 (a) Public Access to Results of Federally Funded Research at the U.S. Geological Survey: Scholarly Publications and Digital Data. Available at http://www2.usgs.gov/quality_integrity/open_access/downloads/USGS-PublicAccessPlan-APPROVED-v1.03.pdf [Last accessed 9 May 2016].

U.S. Geological Survey 2016 (b) USGS Website. Available at <https://www.usgs.gov/about/organization> [Last accessed 9 May 2016].